Swapping Autoencoder for Deep Image Manipulation

Abstract

Deep generative models have become increasingly effective at producing realistic images from randomly sampled seeds, but using such models for controllable manipulation of existing images remains challenging. We propose the Swapping Autoencoder, a deep model designed specifically for image manipulation, rather than random sampling. The key idea is to encode an image with two independent components and enforce that any swapped combination maps to a realistic image. In particular, we encourage the components to represent structure and texture, by enforcing one component to encode co-occurrent patch statistics across different parts of an image. As our method is trained with an encoder, finding the latent codes for a new input image becomes trivial, rather than cumbersome. As a result, it can be used to manipulate real input images in various ways, including texture swapping, local and global editing, and latent code vector arithmetic. Experiments on multiple datasets show that our model produces better results and is substantially more efficient compared to recent generative models.